SUPPLEMENTARY MATERIALS

THE BOX COVERING METHOD

Since the box covering method is central to the understanding of the scale-invariant properties of networks, we describe it in more detail here. Figure 1a shows the same network as in Fig. 1a in the manuscript for the case $\ell_B = 2$. We tile the system by first assigning nodes 1 and 2 to the box colored in blue. Notice that the maximum distance between the nodes of a given box is $\ell_B - 1$. Thus, node 8 would not be in the blue box since its distance from node 2 is $\ell = 2$ (even though its distance from 1 is $\ell = 1$). Then we cover the nodes 6 and 7 with the orange box, and the nodes 3, 4, and 5 with the red box. Finally, the last node 8 is assigned to the green box. The number of boxes to cover the network is then $N_B = 4$.

The renormalization is then applied by replacing each box by a single node. Thus, nodes 1 and 2 will be combined into a single node as indicated by the arrow from the first panel to the second panel in Fig. 1a. This renormalized node is connected with the orange and green boxes because there is a link between nodes 2 and 7, and 1 and 8, respectively. The same rule applies to the other boxes. The renormalized network is shown in the second panel. The system is then tiled again with boxes; in this case two boxes (blue and red) are needed to cover the entire network. The two boxes are then replaced by nodes and a second renormalized network is obtained as shown in the third panel. Finally, the last two nodes belong to the same (red) box and are replaced by a single node.

This procedure is applied to the WWW in Fig. 1b in the manuscript. The main panel corresponds to the first stage in the renormalization of the web for $\ell_B = 3$. The procedure is applied again obtaining the remaining panels in Fig. 1b in the manuscript until the web is reduced to a single box in the last panel. The colors of the nodes corresponds to the boxes to which they belong.

Figure 2d in the manuscript shows the invariance of the degree distribution $P(k)$ under the renormalization performed as a function of the box size in the WWW. The other networks analyzed in this study present the same invariant property. It is important to mention that the networks are also invariant under multiple renormalizations applied for a fixed box size ℓ_B. This corresponds, for instance, to the stages depicted in Fig. 1a in the manuscript in the

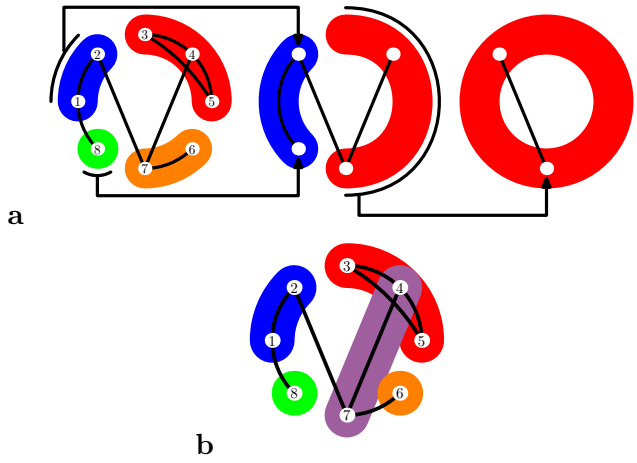


FIG. 1: Details of the box covering method for **a,** $\ell_B = 2$. **b,** A different covering for the same network as in (a) for $\ell_B = 2$. Different coverings give raise to the same exponents as explained in the text.

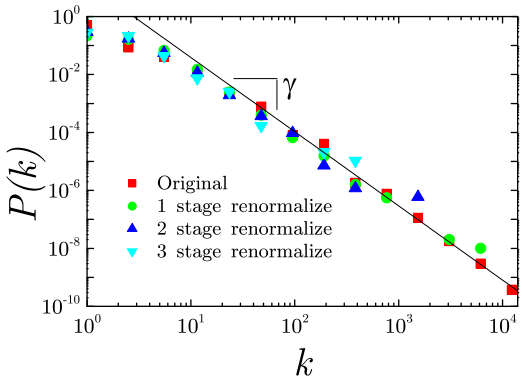


FIG. 2: Invariance of the degree distribution of the WWW under multiple renormalizations done at fixed $\ell_B = 3$. The stages 1, 2, and 3 correspond to the networks depicted in the first three stages in Fig. 1b in the manuscript.

second row for $\ell_B = 3$ for the network demo. Figure 2 shows the invariance of $P(k)$ for the WWW after several stages of the renormalization for a fixed $\ell_B = 3$, and it is the analogous of Fig. 2d in the manuscript for different box size. The stages 1, 2, and 3 correspond to the networks depicted in the first 3 stages in Fig. 1b in the manuscript.

From the above explanation it should be clear that there are many ways to tile the network. For instance in Fig. 1b we show another tiling. In this case we assign nodes 4

and 7 together in a single box instead of nodes 6 and 7 as in Fig. 1a. This tiling results in an extra box needed to cover node 6 and therefore in a larger number of nodes to tile the system, $N_B = 5$.

While there are many ways to assign nodes to the boxes, we notice that the rigorous mathematical definition of Eq. (3) corresponds to the *minimum* number of boxes needed to cover the network. This minimization does not have any consequence for the determination of the fractal dimension in homogeneous clusters. However, it may become relevant when calculating the self-similar exponent of a complex network with a *widely* distributed number of links. Finding the minimum number of boxes to cover the network is a hard optimization problem to solve, analogous to the graph coloring problem in the NP-complete complexity class. This minimization problem has to be solved by an exhaustive numerical search since there is no numerical algorithm to solve this kind of problems.

We have performed the search over a limited part of the phase-space for the WWW to obtain an estimation of the average and the minimum number of boxes needed to tile the network for every value of ℓ_B. We find that the average value of the boxes is very close to the estimated minimum number of boxes. Moreover, we find that the minimization is not relevant and any covering gives rise to the same exponent.

SCALE-FREE TREE STRUCTURE

The underlying meaning of the existence of scale-free networks which are self-similar is yet to be deciphered, but some insight can be gained by examining the simplest structure of a known network of that kind: a *tree* network which has been characterized using field theoretical arguments and fractal dimensions in [1].

The sequence of renormalization steps depicted in Fig. 1 in the manuscript suggests the following scheme: one begins with a single node and then constructs the network by applying the renormalization procedure in a reversed fashion. This can be achieved by following the procedure in Fig. 1a in the manuscript for a specific value of ℓ_B.

More specifically, a single node with a large number of links is first connected to the next generation of nodes. For every node we assign a number of links from a power-law distribution with a given γ. The next layer of the tree is generated in the same way. A tree structure with a power-law degree distribution and self-similar topology emerges which is

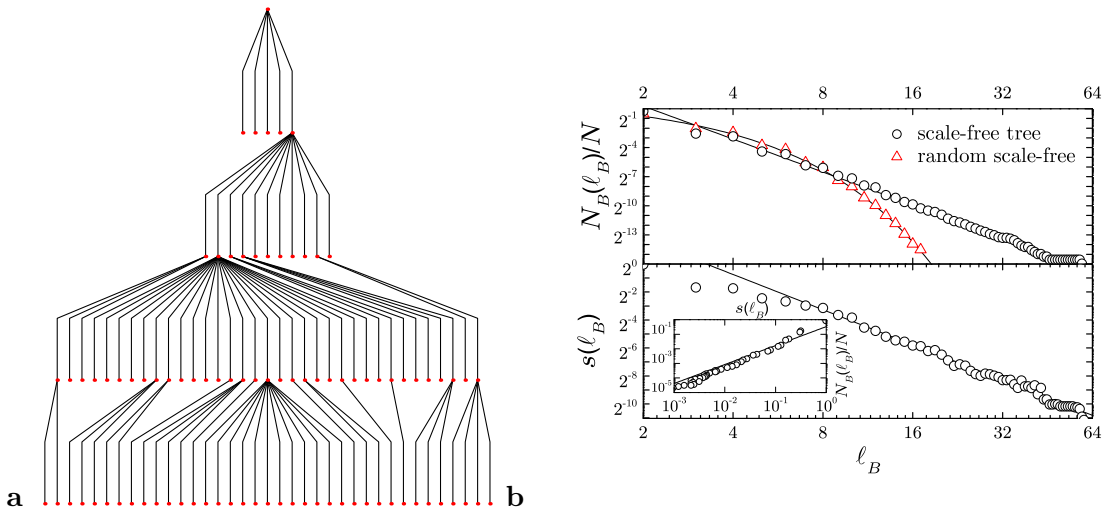
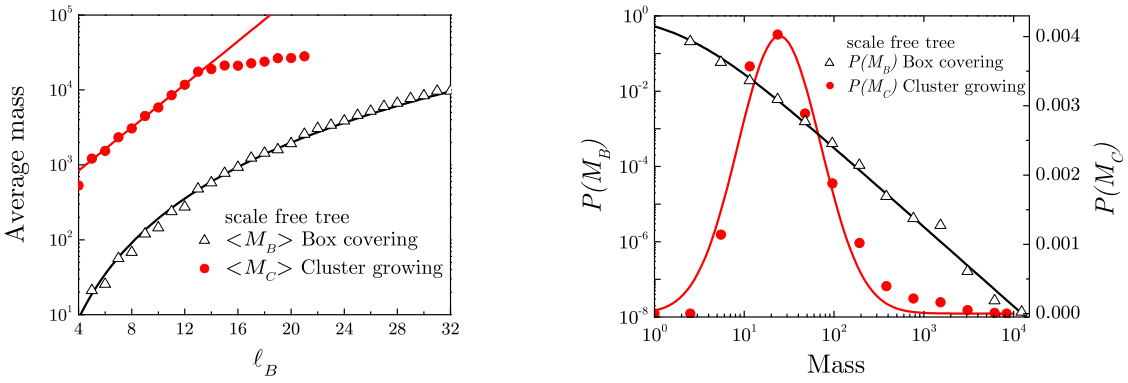


FIG. 3: The scale-free tree structure and the random scale-free model. **a,** Example of a scale-free tree structure. Nodes with a power-law degree distribution are connected in a tree structure without loops. **b,** The log-log plot of N_B vs ℓ_B reveals a self-similar structure for the scale-free tree (upper panel) while $s(\ell_B)$ scales as in Eq. (9) (lower panel). In contrast the random scale-free network where nodes (with a power-law distribution of links) are connected at random shows a lack of self-similarity expressed in the exponential decrease with ℓ_B in the upper panel.

depicted in Fig. 3a.

This is corroborated numerically in Fig. 3b where we study a scale-free tree structure with 192,827 nodes and $\lambda = 2.3$, and we find $d_B = 3.4$ and $d_k = 2.5$. The parallels between the features of such a simple structured network and those discussed in this paper suggest that this simplified view may lie at the core of more complex self-similar networks.

Moreover, we also calculate the average mass of the boxes and the mass of the clusters in the box covering method and the cluster covering method, respectively, and we find the power law of Eq. (5) and the exponential behaviour of Eq. (6) (see Fig. 4a) in agreement with the results of the real networks analyzed in the main manuscript, Fig. 3a. Figure 4b shows the probability distribution of M_B (power-law) and M_c (log-normal) in agreement with previous results as well, Fig. 3b in the manuscript.



a **b**

FIG. 4: Results for the scale-free tree model. **a,** Mean value of the box mass in the box counting, $\langle M_B \rangle$, and mean value of the cluster mass in the cluster growing method, $\langle M_c \rangle$ versus ℓ_B. **b,** Probability distribution of M_B and M_c for $\ell_B = 5$. The results are in agreement with the finding of real networks in Fig. 3 in the manuscript. A power-law distribution is found for M_B while a log-normal distribution is found for M_c as shown by the fits.

INTERNET

It is interesting to note that not all complex networks show the clear self-similarity of the networks presented so far. We analyze the Internet composed of computers and routers linked by physical lines such as the database collected by the SCAN project (the "Mbone", www.isi.edu/scan/scan.html, we also analyze the database of the Internet Mapping Project [2] and found similar results). Figure 5 shows the result of $N_B(\ell_B)$. We fit the curve with a modified power-law

$$N_B(\ell_B) \sim (\ell_B + \ell_c)^{-d_B}, \tag{1}$$

with $\ell_c = 14.9$ representing a cut-off and $d_B = 8.5$, suggesting a large self-similar exponent. The decay of N_B with ℓ_B is faster than a power-law and slower than exponential as shown in the inset of Fig. 5.

Thus these networks lack the clear self-similar structure found for the WWW, actors and the biological networks. However, we find that the distribution of $P(M_B)$ remains a power law and the degree distribution $P(k)$ is invariant under the renormalization suggesting that some self-similar properties might still be valid for the Internet. We notice that Internet maps are made by programs that use the IP protocol to trace the connections between each

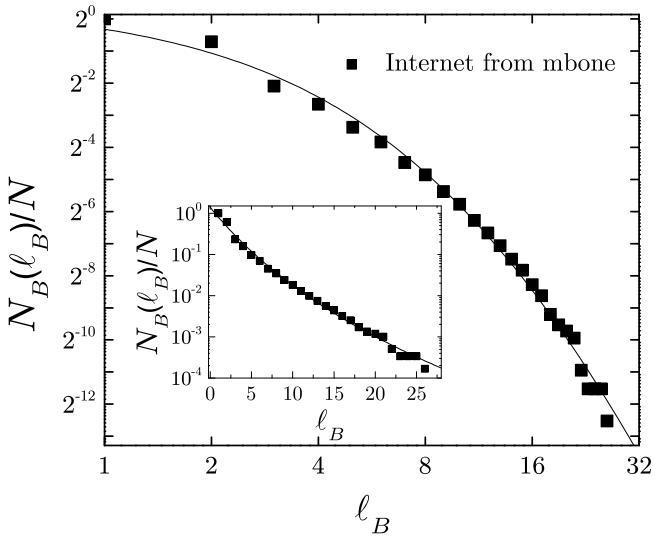


FIG. 5: Internet. Log-log plot of $N_B(\ell_B)$. The solid line represents the modified power law fit, Eq. (1). The inset shows a linear-log plot indicating that the decay is slower than exponential.

registered node in the Internet. These maps are incomplete since they map a few routers from each domain and also due to the existence of firewalls. Thus, the apparent lack of self-similarity might be due to incomplete information of the network.

PROTEIN-PROTEIN INTERACTION NETWORKS

We also analyze the protein interaction networks of the fruit fly *D. melanogaster* as given in [3], the bacterium *H. pylori* [4], the baker's yeast *S. cerevisiae* [5], and the nematode worm *C. elegans* [6], which are all available via the DIP database [7]. Figure 6 shows the results of N_B versus ℓ_B indicating that their behaviour is in between a pure power-law decay and a pure exponential. As with the Internet data, we are able to fit the results with Eq. (1) with $\ell_c = 7.2$ and $d_B = 7.6$ for *C. elegans*. For *H. pylori* and *D. melanogaster* the fit is a pure exponential $N_B(\ell_B) \sim \exp(-\ell_B/\ell_e)$ with $\ell_e \approx 1$, while for *S. cerevisiae* the data could be fitted either by an exponential or by large values of ℓ_c and d_B (note that the exponential is the limit of Eq. (1) for $\ell_c \to \infty$, $d_B \to \infty$ and $\ell_c/d_B =$ constant). On the other hand, we observe that for small scales, N_B seems to display the same power law found for *E. coli* and *H. sapiens*. The lack of clear self-similarity in these networks might be due

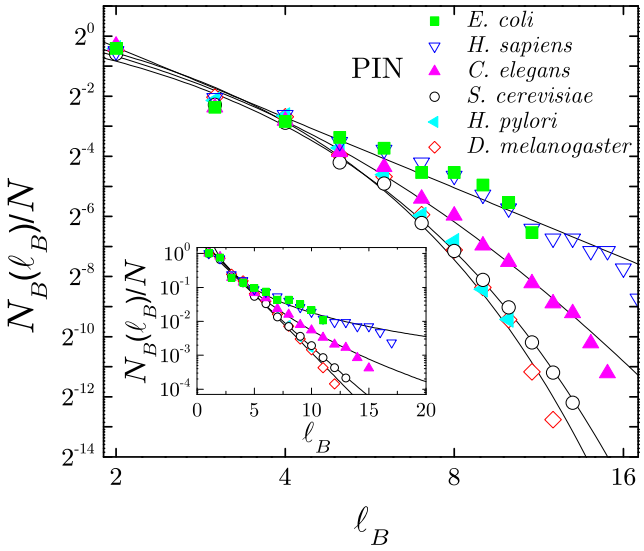


FIG. 6: Scaling for the protein-protein interaction networks. Log-log plot of N_B versus ℓ_B for different protein-protein interaction networks. While *E. coli* and *H. sapiens* show a clear power law behavior, the other protein networks show a modified power-law behaviour or a pure exponential decay. The inset shows a linear-log plot of $N_B(\ell_B)$.

to the incompleteness of these databases which are continuously being updated with newly discovered physical interactions [8].

RANDOM SCALE-FREE NETWORK

Next we introduce an example of a model lacking self-similarity: the random scale-free model. This model consists of nodes to which a number of links are assigned with a power-law degree distribution and then connected randomly. Such a network shows a small world effect and a scale-free property but is not self-similar. We numerically find that the number of boxes decays exponentially with the box size (see Fig. 3b). Moreover, while Eq. (8) is still valid in this case, the power law relation in Eq. (9) is replaced by an exponential law. We conjecture that the reason for this is a clustering of hubs; by assigning randomly the connections between the nodes, two nodes with a large number of links will have a large probability to be connected. This induces spatial correlations in the values of k which may

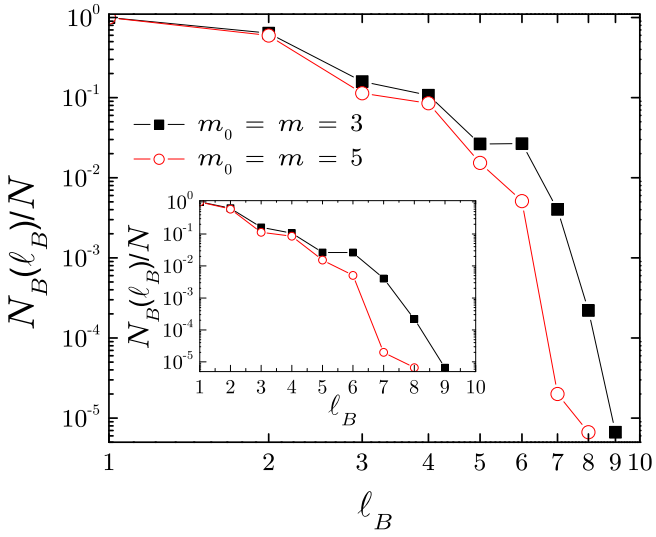


FIG. 7: Barabási-Albert model of scale-free networks with preferential attachment for 150,000 nodes and $m = m_0 = 3$ and $m = m_0 = 5$. m_0 is the initial number of nodes in the system and m is the number of links of a newly created node in the dynamical growth of the network [9]. Log-log plot of N_B versus ℓ_B showing the lack of a power law behaviour. The inset shows a linear-log plot indicating that N_B decreases faster than exponential with ℓ_B.

explain the breakdown of self-similarity. In contrast, the simple tree-structure proposed above does not cluster the hubs by construction. A summary of our results is presented in Table I.

THE BARABÁSI-ALBERT MODEL AND THE ERDÖS-RÉNYI RANDOM GRAPH AT CRITICALITY

We also analyzed the Barabási-Albert model of complex networks [9] (which introduces the concepts of preferential attachment to describe the dynamics of scale-free networks). The results of $N_B(\ell_B)$ are shown in Fig. 7 for different parameters in the model (see [9] for details) reveling that the structure is not self-similar; N_B seems to decrease faster than exponential with ℓ_B.

It is interesting to compare our results with the random Erdös-Rényi graph [10, 11] at the critical percolation threshold. In this case the largest cluster has self-similar properties

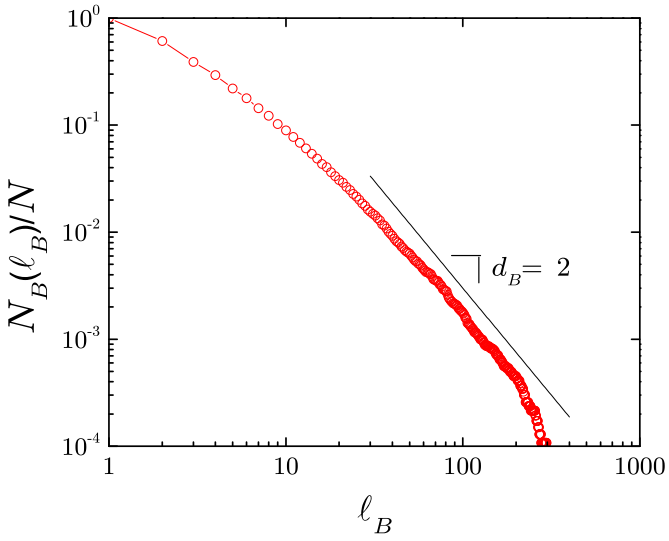


FIG. 8: Erdös-Rényi random graph at criticality. Log-log plot of N_B versus ℓ_B showing the self-similar exponent $d_B = 2$ which is obtained for large distances.

and Eq. (5), $\langle M_B(\ell_B)\rangle \sim \ell_B^{d_B}$, is valid with $d_B = 2$ [12]. We corroborate this result in Fig. 8 showing the scaling of the number of boxes N_B with the box size ℓ_B. However, for this case the network is not small-world since Eq. (6) is not valid— as well as Eq. (1)— but rather the mean distance $\bar{\ell}$ scales as $\langle M_c\rangle^{1/2}$, i.e., a power-law relation rather than the logarithmic relation characteristic of small world networks.

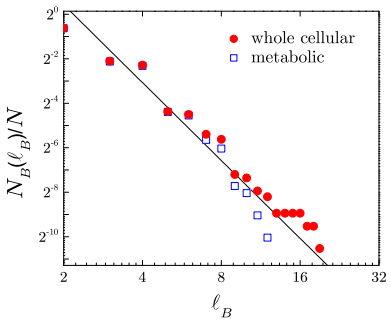
Network	d_B	d_k	$1 + d_B/d_k$ Eq. (10)	γ Eq. (2)
WWW	4.1	2.5	2.6	2.6
Actor	6.3	5.3	2.2	2.2
E. coli (PIN)	2.3	2.1	2.1	2.2
H. sapiens (PIN)	2.3	2.2	2.0	2.1
43 cellular networks	3.5	3.2	2.1	2.2
Scale-free tree	3.4	2.5	2.4	2.3

TABLE I: Summary of the exponents obtained for the scale-invariant networks studied in the manuscript.

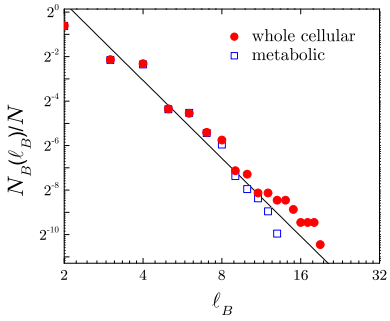
CELLULAR NETWORKS

The WIT database [13] (http://igweb.integratedgenomics.com/IGwit) of cellular networks considers the cellular functions divided according to bioengineering principles containing datasets for intermediate metabolism and bioenergetics (core metabolism), information pathways, electron transport, and transmembrane transport. The metabolic network is a subset of all reactions that take place in the cell. Since this is the largest part of the network we analyze it separately and compare it with the full biochemical reaction network. The data presented in Fig. 2c in the main manuscript represents the full biochemical reaction networks of only three substrates. Here we present results of the 43 different substrates represented in the database for the metabolic and full networks. The following figures show the results of N_B vs ℓ_B. Both the metabolic and full networks display the power law relationship of self-similar networks with the same exponent (within error bars) for all the organisms considered (the metabolic networks show a finite size effect due to their smaller size). We find an average $d_B = 3.5$. The solid line in the figures represent the average fit. The values are reported in Table I.

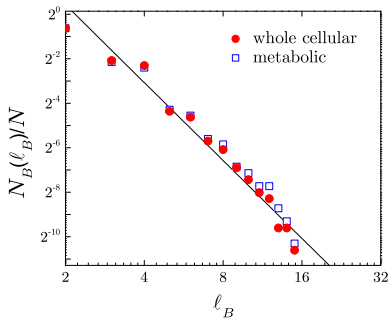
Aquifex aeolicus



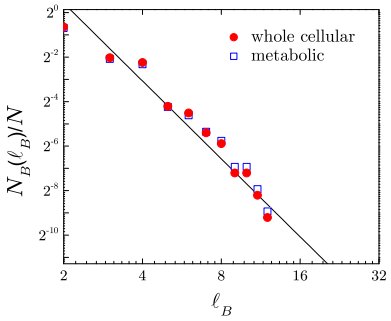
Actinobacillus actinomycetemcomitans



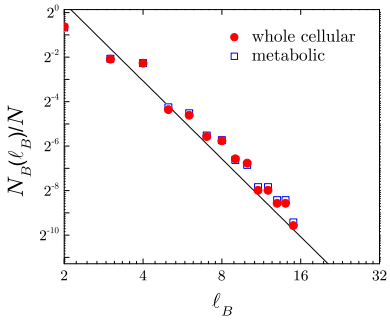
Archaeoglobus fulgidus



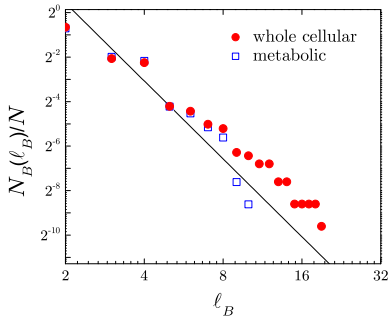
Aeropyrum pernix



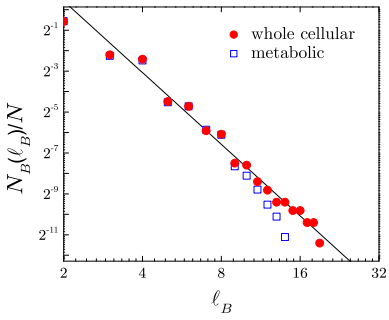
Arabidopsis thaliana



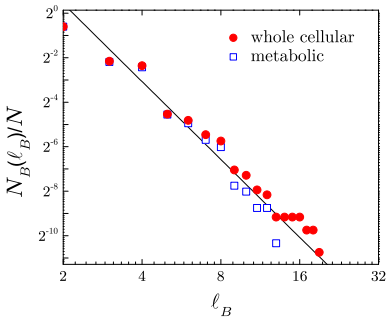
Borrelia burgdorferi



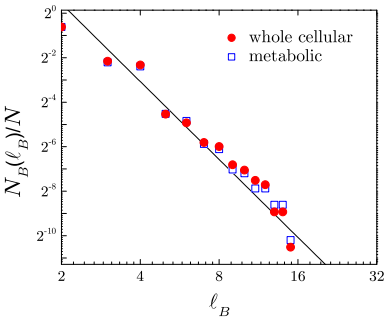
Bacillus subtilis



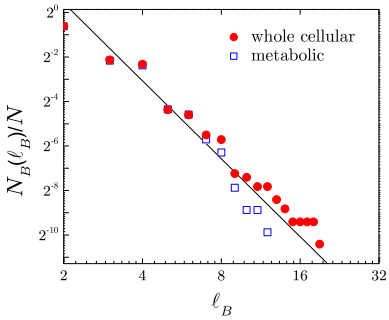
Clostridium acetobutylicum



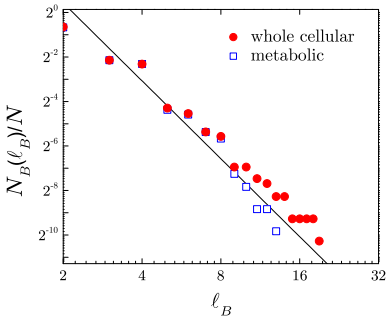
Caenorhabditis elegans



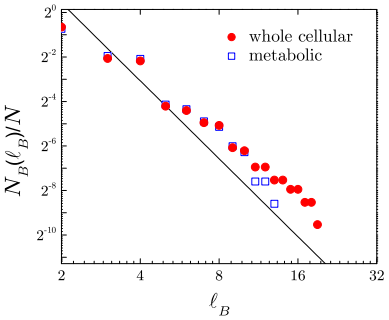
Campylobacter jejuni



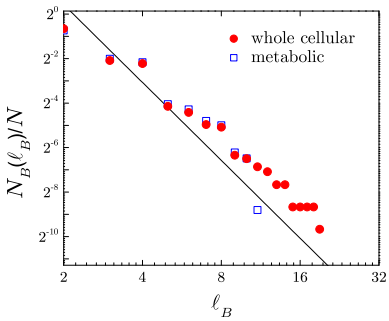
Chlorobium tepidum



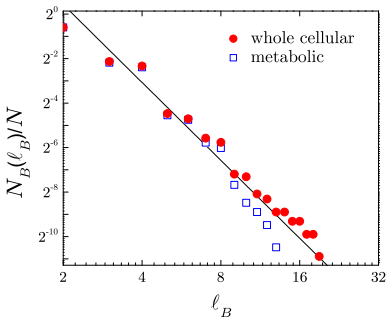
Chlamydia pneumoniae



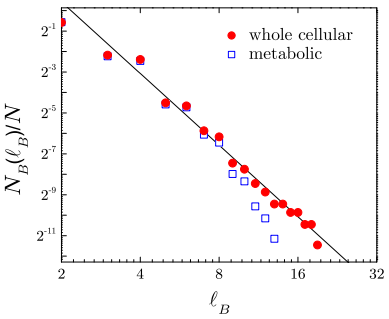
Chlamydia trachomatis



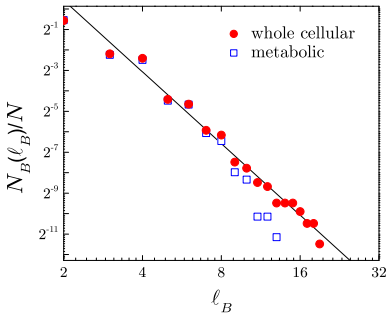
Synechocystis sp.



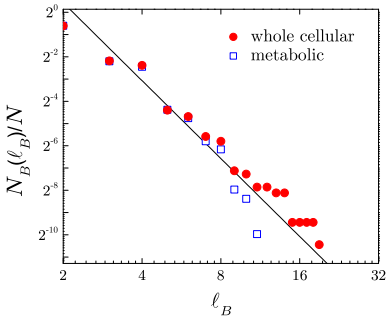
Deinococcus radiodurans



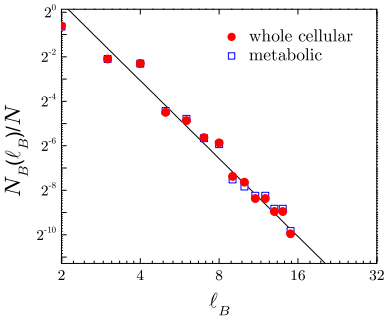
Escherichia coli



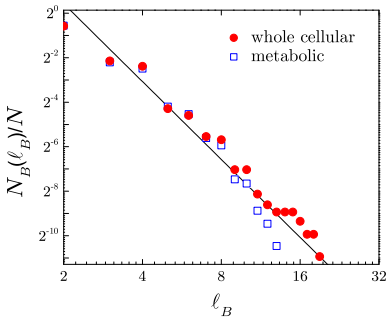
Enterococcus faecalis



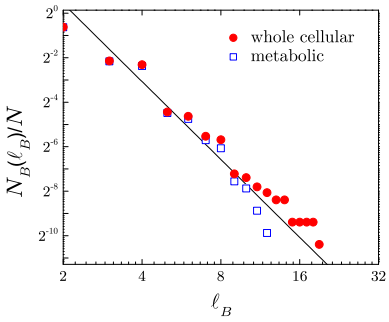
Emericella nidulans



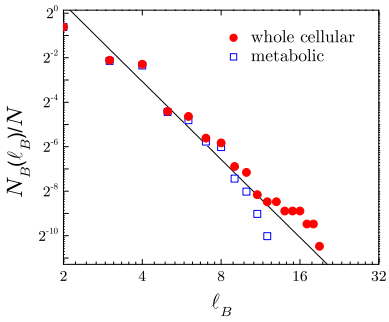
Haemophilus influenzae



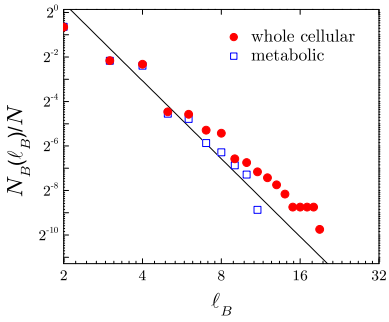
Helicobacter pylori



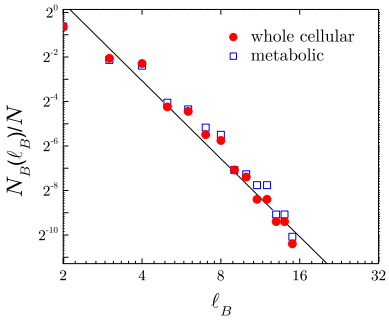
Mycobacterium bovis



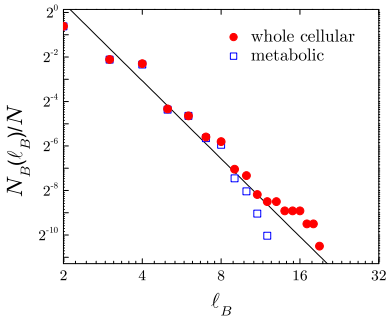
Mycoplasma genitalium



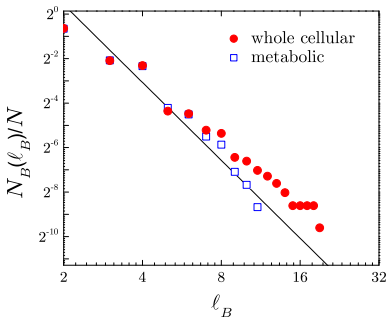
Methanococcus jannaschii



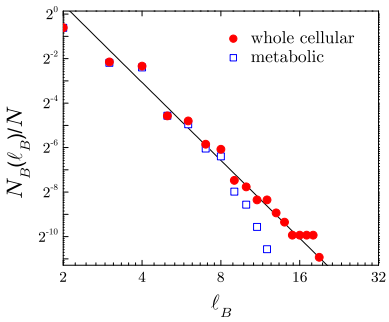
Mycobacterium leprae



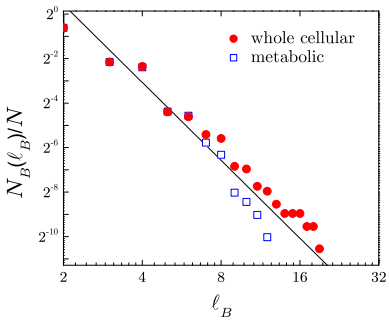
Mycoplasma pneumoniae



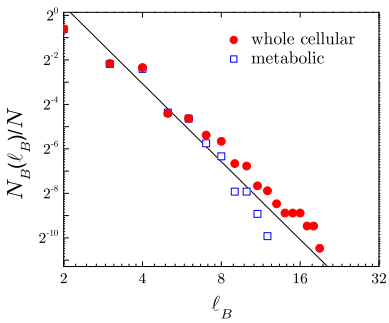
Mycobacterium tuberculosis



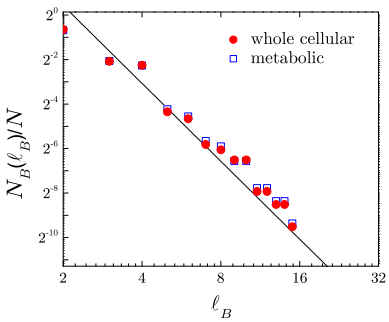
Neisseria gonorrhoeae



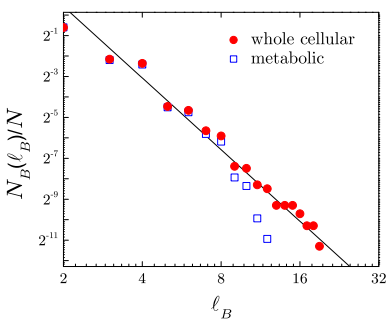
Neisseria meningitidis



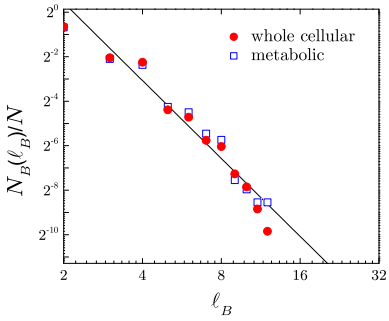
Oryza sativa



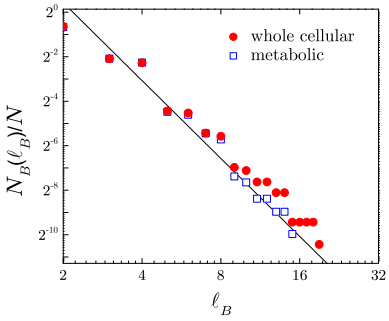
Pseudomonas aeruginosa



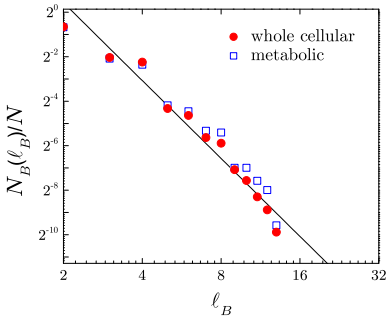
Pyrococcus furiosus



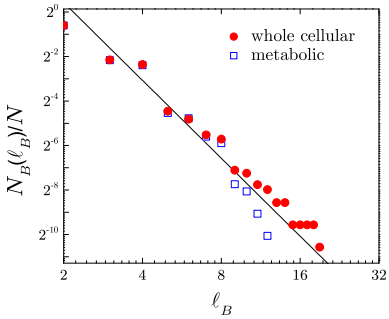
Porphyromonas gingivalis



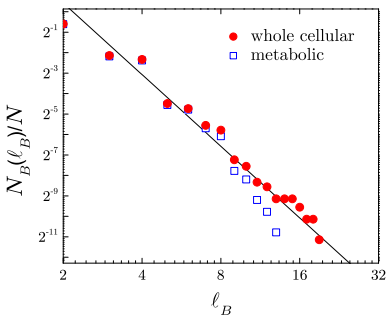
Pyrococcus horikoshii



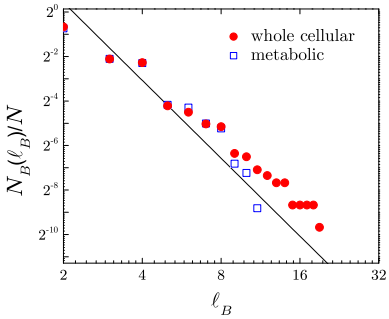
Streptococcus pneumoniae



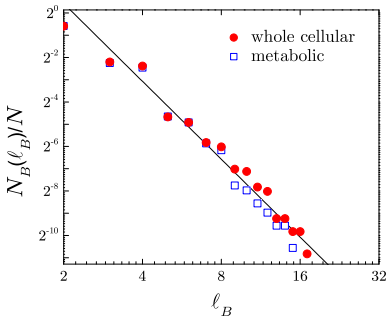
Rhodobacter capsulatus



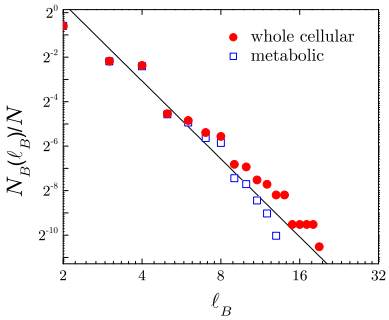
Rickettsia prowazekii



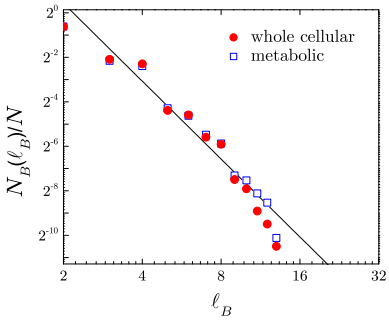
Saccharomyces cerevisiae



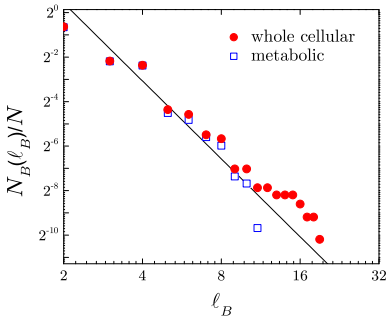
Streptococcus pyogenes



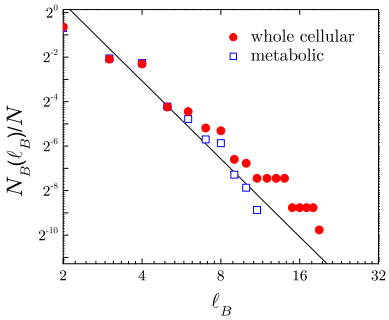
Methanobacterium thermoautotrophicum



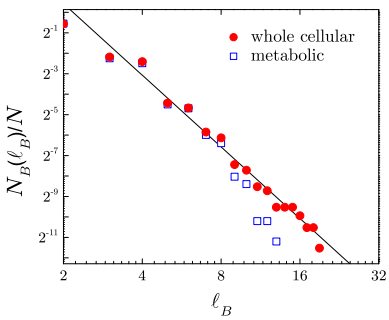
Thermotoga maritima



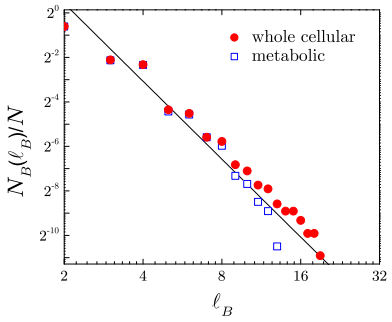
Treponema pallidum



Salmonella typhi



Yersinia pestis



References

[1] Burda, Z., Correia, J. D. & Krzywicki, A. Statistical ensemble of scale-free random graphs *Phys. Rev. E* **64**, 046118 (2001).

[2] Burch, H. & Cheswick, W. Mapping and Visualizing the Internet. *IEEE Computer* **32**, 4 (1999). http://research.lumeta.com/ches/map

[3] Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster. Science* **302**, 1727-1736 (2003).

[4] Rain, J.-C. *et al.* The protein-protein interaction map of *Helicobacter pylori. Nature* **409**, 211-215 (2001).

[5] Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae. Nature* **403**, 623-627 (2000).

[6] Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans. Science* **303**, 540-543 (2004).

[7] *Database of Interacting Proteins* (DIP). http://dip.doe-mbi.ucla.edu

[8] Xenarios, I. *et al.* DIP: the database of interacting proteins. *Nucleic Acids Res.* **28**, 289-291 (2000).

[9] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509-512 (1999).

[10] Erdös, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17-61 (1960).

[11] Bollobás, B. *Random Graphs* (Academic Press, London, 1985).

[12] Braunstein, L., Buldyrev, S. V., Cohen, R., Havlin, S. & Stanley, H. E. Optimal paths in disordered complex networks. *Phys. Rev. Lett.* **91**, 168701 (2003).

[13] Jeong, H, Tombor, B., Albert, R., Oltvai Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651-654 (2000).